frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 18 July 2023
DOI 10.3389/fpubh.2023.1151821

OPEN ACCESS

EDITED BY
Monica Trif,
Centre for Innovative Process Engineering,
Germany

REVIEWED BY
Xiangzhu Zhu,
Vanderbilt University, United States
Drew Day,
Seattle Children's Research Institute,
United States

*CORRESPONDENCE
Stefano Renzetti
✉ stefano.renzetti@unibs.it

RECEIVED 01 February 2023
ACCEPTED 02 June 2023
PUBLISHED 18 July 2023

CITATION
Renzetti S, Gennings C and Calza S (2023) A weighted quantile sum regression with penalized weights and two indices.
Front. Public Health 11:1151821.
doi: 10.3389/fpubh.2023.1151821

A weighted quantile sum regression with penalized weights and two indices

Stefano Renzetti[1]*, Chris Gennings[2] and Stefano Calza[3]

[1]Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, Università degli Studi di Brescia, Brescia, Italy, [2]Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [3]Department of Molecular and Translational Medicine, Università degli Studi di Brescia, Brescia, Italy

Background: New statistical methodologies were developed in the last decade to face the challenges of estimating the effects of exposure to multiple chemicals. Weighted Quantile Sum (WQS) regression is a recent statistical method that allows estimating a mixture effect associated with a specific health effect and identifying the components that characterize the mixture effect.

Objectives: In this study, we propose an extension of WQS regression that estimates two mixture effects of chemicals on a health outcome in the same model through the inclusion of two indices, one in the positive direction and one in the negative direction, with the introduction of a penalization term.

Methods: To evaluate the performance of this new model we performed both a simulation study and a real case study where we assessed the effects of nutrients on obesity among adults using the National Health and Nutrition Examination Survey (NHANES) data.

Results: The method showed good performance in estimating both the regression parameter and the weights associated with the single elements when the penalized term was set equal to the magnitude of the Akaike information criterion of the unpenalized WQS regression. The two indices further helped to give a better estimate of the parameters [Positive direction Median Error (PME): 0.022; Negative direction Median Error (NME): −0.044] compared to the standard WQS without the penalization term (PME: −0.227; NME: 0.215). In the case study, WQS with two indices was able to find a significant effect of nutrients on obesity in both directions identifying sodium and magnesium as the main actors in the positive and negative association, respectively.

Discussion: Through this work, we introduced an extension of WQS regression that improved the accuracy of the parameter estimates when considering a mixture of elements that can have both a protective and a harmful effect on the outcome; and the advantage of adding a penalization term when estimating the weights.

KEYWORDS

environmental mixture, weighted quantile sum regression, two indices, penalized weights, nutrients, obesity

Introduction

Humans are exposed to many chemicals from multiple chemical classes, based on biomonitoring data, which may influence a particular disease or health state (1–3). Of particular concern is that many of these exposures may act jointly (e.g., along a common adverse pathway) so that even low levels of each may together have an adverse effect. Not accounting for this *mixture effect* may underestimate the potential health risk. Analogously, the food we eat includes many important nutrients. Evaluating dietary quality based on a single nutrient is not adequate.

Several new statistical methodologies were developed to face the challenges of estimating the effects of exposure to multiple chemicals, each addressing its own research question (4). These new methods were developed to solve problems like multiple comparisons and multicollinearity that are commonly encountered with high dimensional and correlated exposures. When using classical statistical models, an increased probability of incurring false positive or false negative results can occur in the first case (5–9) while multicollinearity can produce unstable and biased parameter and standard error estimates in the second case (3, 4, 10–18).

Weighted Quantile Sum (WQS) regression is a recent statistical method that is increasingly applied in epidemiological studies to address the research questions of (i) is there a mixture effect associated with a specific developmental or health effect; and (ii) which of the measured components characterize the mixture effect. This method builds an empirically weighted index that represents the mixture effect in an ensemble first step and tests its association with the outcome of interest in a second step. This body burden index reduces the dimensionality and is more robust to multicollinearity (19, 20). The original methodology provides estimation of a single empirically weighted index that measures the association between the mixture and the dependent variable in only one direction (either positive or negative), which may be interpreted as the joint action of the components in that direction (i.e., a mixture effect). On the other hand, other methods (e.g., variable selection methods) focus on parsimony in predicting an outcome and thus address different research questions. Further, shrinkage methods suffer from limitations in the presence of highly correlated variables like the grouping effect in the elastic net and the arbitrary selection of variables in the LASSO that can be problematic in the risk evaluation of environmental mixtures (19). The estimation of a single index can be an advantage when the elements in the mixture have the same direction in the association with the dependent variable as it focuses inference in the important direction, as is the case when evaluating a mixture effect of jointly acting components. In fact, looking in one direction we avoid the reversal paradox (18) and we increase the power to detect a mixture effect using a single degree of freedom test. However, when the mixture is made of both "good" and "bad" actors (i.e., two sets of components where one set is related to a positive direction and opposite for the other set), it can become a limitation when we want to estimate both the positive mixture effect and negative mixture effect on the specific outcome of interest in a single analysis.

In this study, we propose an extension of WQS regression where two indices are estimated [termed two-indices WQS (2iWQS) in the sequel], one in the positive and the other in the negative direction, in the same model both at the nonlinear estimation ensemble step where the weights are determined and at the final model inference step. This will allow us to use a data reduction approach trying to focus the inference for joint action in now both directions still based on the idea of a mixture effect with a single degree of freedom test for each direction. The simultaneous estimation of the two indices is made possible through the addition of a penalization term when estimating the weights. An application of the new method uses National Health and Nutrition Examination Survey (NHANES) (2011–2016) dietary data and association with obesity. A simulation study is also presented to evaluate the performance of the method.

Methods

The general formula for the WQS generalized nonlinear regression model for *c* components in the mixture is the following:

$$g(\mu) = \beta_0 + \beta_1\left(\sum_{i=1}^{c} w_i q_i\right) + \mathbf{z}'\boldsymbol{\phi}$$

where w_i are the unknown weights associated to each component of the mixture to be estimated through an ensemble step using bootstrap samples (19) or random subset samples (21), q_i are the values of the components scored into quantiles (quartiles, deciles,…), β_0 is the intercept, β_1 is the coefficient associated to the WQS index, $\mathbf{z}'\boldsymbol{\phi}$ are the vector of covariates and parameters, respectively and $g()$ is any link function between the mean μ and the predictor variables as in generalized linear models. WQS regression requires that data are split in a training and validation dataset. The first part of the data is used for the ensemble step to determine the weighted index while accommodating for the correlation among the components, and the final model is fitted on the holdout data to test the significance of β_1. For the estimated weights the following constraints are imposed: $\sum_{i=1}^{c} w_i = 1$; and $0 \le w_i \le 1$.

Once the weights are estimated for each ensemble step sample the WQS index is estimated through the formula: $WQS = \sum_{i=1}^{c} \overline{w}_i q_i$; where $\overline{w}_i$ is the mean of the weights found in the ensemble steps associated to either a positive or a negative $\hat{\beta}_1$ depending on the chosen direction of the association between the mixture and the outcome:

$$\overline{w}_i = \frac{1}{\sum_{b=1}^{B} f\left(\hat{\beta}_{1(b)}\right)} \sum_{b=1}^{B} \mathrm{w}_{i(b)} f\left(\hat{\beta}_{1(b)}\right)$$

where $f\left(\hat{\beta}_{1(b)}\right)$ is a signal function defined, for example, as the inverse of the absolute value of the t-test statistic for $\hat{\beta}_1$, or t^2, or e^t (22). The final model is fitted on the holdout validation dataset using the generalized linear model $g(\mu) = \beta_0 + \beta_1 WQS + \mathbf{z}'\boldsymbol{\phi}$.

In this study we propose to first include a penalized weight estimate to better identify the truly associated elements in the case of highly correlated data. The objective function (with an identify link) in this case is as follow:

$$\hat{\theta} = \underset{\theta}{\arg\min}\left\{\sum_{j=1}^{n}\left(y_j - \left(\beta_0 + \beta_1\left(\sum_{i=1}^{c} w_i q_i\right) + \mathbf{z}'\boldsymbol{\phi}\right)\right)^2 + \lambda\sum_{i=1}^{c}\left|v_i\right|\right\}$$

where $\theta = \left(\beta_0, \beta_1, v_1, \ldots, v_c, \boldsymbol{\phi}, \lambda\right)$ and where we defined $w_1 = \frac{v_i^2}{\sum_{i=1}^{c} v_i^2}, v_i \in \Re, i = 1, \ldots, c$ to make the penalization term effective.

We further introduce two indices in the same model to allow an estimate of the mixture effect in a positive direction and a mixture effect in a negative direction at the same time, both in the training and validation steps. The new general formula is the following:

$$g\left(\mu\right) = \beta_0 + \beta_{1p}\left(\sum_{i=1}^{c} w_{pi} q_i\right) + \beta_{1n}\left(\sum_{i=1}^{c} w_{ni} q_i\right) + \mathbf{z}'\boldsymbol{\phi}$$

where w_{pi} and w_{ni} are the unknown weights associated with each component for the positive and negative direction, respectively, while β_{1p} and β_{1n} are the two parameters that measure the positive and negative effect of the mixture on the outcome. The two indices will be kept in the model both in the first step where two set of weights are estimated (one for the positive and one for the negative direction) and in the second step when the final model is fitted. The equality and inequality constraints are applied to both sets of weights besides a constraint to each β_{1j} parameter: $\beta_{1p} \geq 0$ and $\beta_{1n} \leq 0$. The penalization term is also considered to better discriminate between the elements having an effect and those not associated with the outcome and to reduce the noise produced by the null components that can increase the correlation between the two indices. Without loss of generality, the objective function for the identity link is of the form:

$$\hat{\theta} = \underset{\theta}{\arg\min}\left\{\sum_{j=1}^{n}\left(y_j - \left(\beta_0 + \beta_{1p}\left(\sum_{i=1}^{c} w_{pi} q_i\right) + \beta_{1n}\left(\sum_{i=1}^{c} w_{ni} q_i\right) + \mathbf{z}'\boldsymbol{\phi}\right)\right)^2 + \lambda\left(\sum_{i=1}^{c}\left|v_{pi}\right| + \sum_{i=1}^{c}\left|v_{ni}\right|\right)\right\}$$

where $\theta = \left(\beta_0, \beta_{1p}, \beta_{1n}, v_{p1}, \ldots, v_{pc}, v_{n1}, \ldots, v_{nc}, \boldsymbol{\phi}, \lambda\right)$ and

$$w_{di} = \frac{v_{di}^2}{\sum_{i=1}^{c} v_{di}^2}, v_{di} \in \Re, d = \mathrm{p}, n, i = 1, \ldots, c.$$

To set the starting values of the v_{di} we fit two standard WQS regressions constraining the beta in each direction and considering all the observations without splitting the dataset in training and validation and without bootstrapping (for these two starting models we initialize the v_{di} at 1). We then extract the estimated v_{di} from the two models and use them as starting values for the training step of the 2iWQS regression. If one of the two initial WQS regressions does not converge and we are not able to find a set of starting values for the weights in one of the two directions we propose to set the v_{di} equal to the inverse of the elements included in the mixture ($v_{di} = \frac{1}{c}$).

The final model for the validation step is $g\left(\mu\right) = \beta_0 + \beta_{1p} WQS_p + \beta_{1n} WQS_n + \mathbf{z}'\boldsymbol{\phi}$.

To further control the collinearity between the two indices we apply a different signal function when averaging the weights estimated in each ensemble sample in the training step. In particular, we considered the tolerance (the inverse of the variance inflation factor (VIF)) as the weight in the weighted mean: $f\left(\hat{\beta}_{1d(b)}\right) = \left(tol\left(\hat{\beta}_{1d(b)}\right) / \sum_{b=1}^{B} tol\left(\hat{\beta}_{1d(b)}\right)\right)^k$ where tol is the tolerance associated to the parameter $\hat{\beta}_{1d(b)}$, j refers to the direction ($d = p$ for positive direction, $d = n$ for negative direction) and k is chosen depending on the variability of the tolerance values: higher k is applied with lower variability to better discriminate between those models where there is higher collinearity between the positive and the negative index and those where collinearity is less severe. For example, we start with $k = 2$; if the VIF for either WQS parameter exceeds 5 then we may increase k to, say, 3. In the following simulation and case studies we set $k = 3$ since it allowed to find two indices with a VIF below 5.

To define the shrinkage parameter, lambda, a cross-validation step should be performed. However, since this can be computationally intense, a rule of thumb that can be applied to choose the value of the parameter lambda is to set it equal or close to the magnitude of the Akaike Information Criterion (AIC) of the non-penalized WQS regression. An outline of the steps to take when trying to fit a WQS regression with a single or double index follows:

- Step 1: fit a standard WQS regression (with single or double index as needed)
- Step 2: set three different shrinkage parameter values, one equal to the magnitude of the AIC of the regression fitted at step 1 one to a lower and one to a greater order of magnitude and follow a rough bisection algorithm. Alternatively, the search of the best lambda can also be refined by looking to the intermediate values, e.g., if the magnitude of the AIC is 1,000, then lambda can be set equal to 500 and 5,000 besides 100 and 10,000.
- Step 3: In the case of the 2iWQS a final check of the correlation between the two indices is needed. The exponent in the signal function can be increased in the case of multicollinearity.

The possibility to fit a 2iWQS and to estimate penalized weights will be added to the gWQS R package (23).

Simulation study

To evaluate the performance of this new model in terms of the estimation of the regression parameters and the weights, we performed a simulation study where we compared the results obtained by the 2iWQS with those obtained by the standard WQS regression and the quantile g-computation, a novel statistical method that combines WQS regression and g-computation to provide an effect estimate per simultaneous quantile increase in all elements (i.e., overall mixture effect), as well as weights that represent the importance of individual components of the mixture in the positive or negative direction (24). For a direct comparison with the directional sum mixture effect estimates (β_{1d}) of WQS regression, we will evaluate the analogous directional sum mixture effect estimates of the quantile g-computation models (ψ_d) rather than the overall mixture effect (ψ). The former are calculated in quantile g-computation models as the sum of all mixture component coefficients in a given direction, while the latter is the sum of all mixture component coefficients (24). In all simulation scenarios, we have applied a repeated holdout approach (25) for all standard WQS regression and 2iWQS regression models to have more power in the parameter estimates and to reduce variability in the weights (26). We set 50 different training and validation splits of the dataset with 50 bootstrap iterations performed on each training set; we used 50 instead of the typical 100 or more to limit computational time. We took the data from the NHANES 2011–2012, 2013–2014, and 2015–2016 survey cycles where a total of 38 nutrients were measured through the administration of a food frequency questionnaire. In total 100 different datasets were built generating the 38 variables from a multivariate normal distribution keeping the same correlation structure of the original data. As we can see from Figure 1 the nutrient data show a complex correlation matrix with Spearman estimates ranging from −0.08 to 0.883.

Based on the case study results (see results section) the corresponding five nutrients with a median weight exceeding the threshold were selected for the positive direction while the corresponding 10 elements were chosen in the negative direction. In particular, the dependent variable was generated from a normal distribution with mean equal to the combination obtained by applying the parameters given in Table 1 to the 2iWQS formula and a unit standard deviation. The case study weights were rescaled assigning a null weight to all the non-selected nutrients. One more scenario was considered halving the values of the correlation matrix. This simulation study is structured such that the sensitivity in the positive direction (i.e., estimating weights exceeding the threshold of 1/38 = 0.026) is particularly difficult when two of the four nutrients have "true" weights of 0.05 and 0.07. As a last scenario, we considered a unidirectional association between the mixture and the outcome and we used only the positive weights to generate the dependent variable as described above.

Case study

We applied the new method of the 2iWQS regression to assess the effects of nutrients on obesity among adults using the NHANES 2011–2016 data. Consent from subjects participating in the study was received

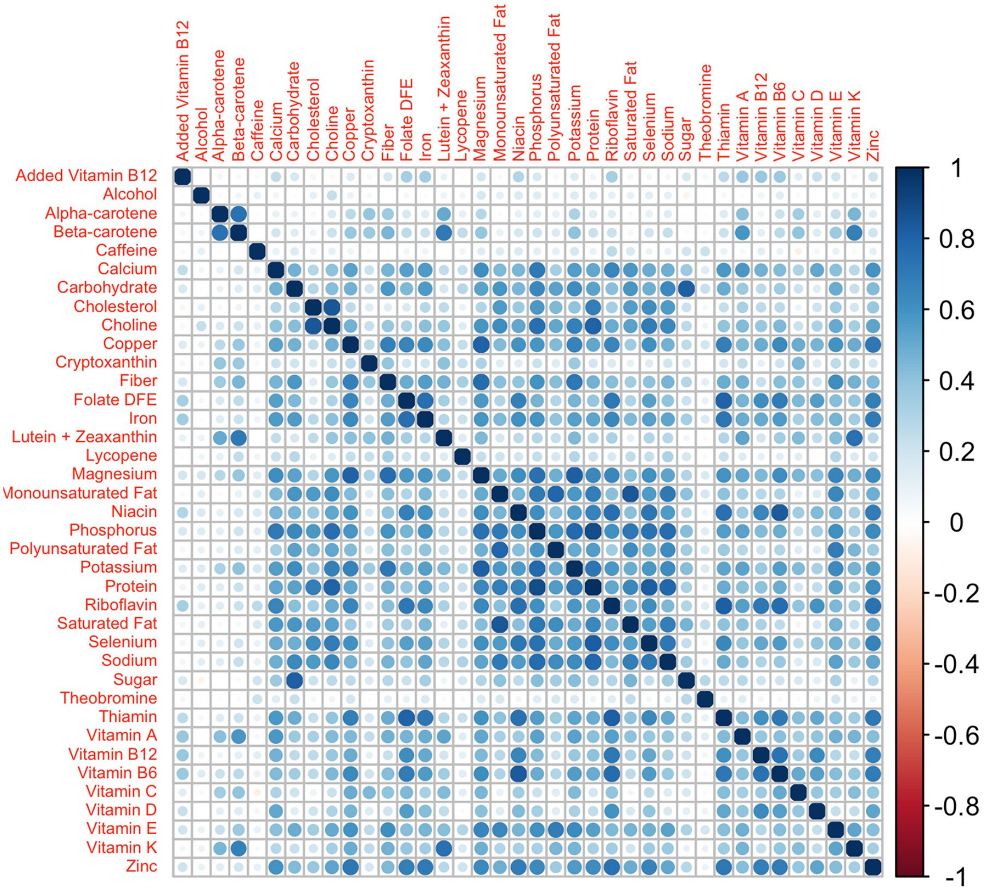

FIGURE 1
Nutrients' correlation matrix: correlation matrix among the 38 nutrients from the NHANES 2011–2012, 2013–2014, and 2015–2016 survey cycles.

TABLE 1 Parameters' value used in the simulation study: values of the parameter regression and weights used to generate the dependent variable.

Parameter	PWQS	NWQS
β_{1p}	0.5	
β_{1n}		−0.5
w_{sodium}	0.50	0
$w_{polyunsaturated\ fat}$	0.19	0
$w_{cholesterol}$	0.15	0
$w_{caffeine}$	0.10	0
$w_{calcium}$	0.07	0
$w_{magnesium}$	0	0.27
w_{fiber}	0	0.13
$w_{vitamin\,C}$	0	0.12
$w_{vitamin\,B6}$	0	0.09
$w_{beta-carotene}$	0	0.08
$w_{folate\,DFE}$	0	0.08
$w_{vitamin\,D}$	0	0.07
$w_{vitamin\,E}$	0	0.06
$w_{vitamin\,K}$	0	0.06
$w_{alpha-carotene}$	0	0.05

The weight of all the remaining variables not included in the table were set to 0.
DFE, Dietary Folate Equivalents; PWQS, Positive Weighted Quantile Sum index; NWQS, Negative Weighted Quantile Sum index.

prior to conducting the study and the study has been reviewed and approved by the CDC/NCHS Ethics Review Board (ERB).

In total 21,798 subjects with reliable dietary data (meaning that all relevant variables associated with the 24-h dietary recall contain a value) were included in the analysis; 2,459 subjects were excluded because they were on any kind of diet to lose weight or for another health-related reason at the time of the interview; 11,926 subjects were younger than 20 or older than 60 years old and 1,453 subjects had a missing value at least for one of the covariates considered in the analysis. In total, 5,960 subjects were included in the study (Figure 2).

Obesity was defined as Body Mass Index (BMI) greater or equal to 30 kg/m^2.

Nutrients were estimated from the dietary intake data that considered the types and amounts of foods and beverages (including all types of water) consumed during the 24-h period prior to the interview (midnight to midnight). Two interviews were performed: the first one was collected in-person while the second interview was collected by telephone 3–10 days later. Details of the survey are described elsewhere (27, 28). In this study, we averaged the two nutrients when both evaluations were considered as usual food consumption compared to the food habits of each participant (only one measurement was included in the analysis if the other one was not usual while the observation was dropped if both evaluations were not usual; this was a self-reported answer to the question of whether the person's overall intake on the previous day was much more than usual, usual, or much less than usual) and we added the dietary supplement intake when applicable.

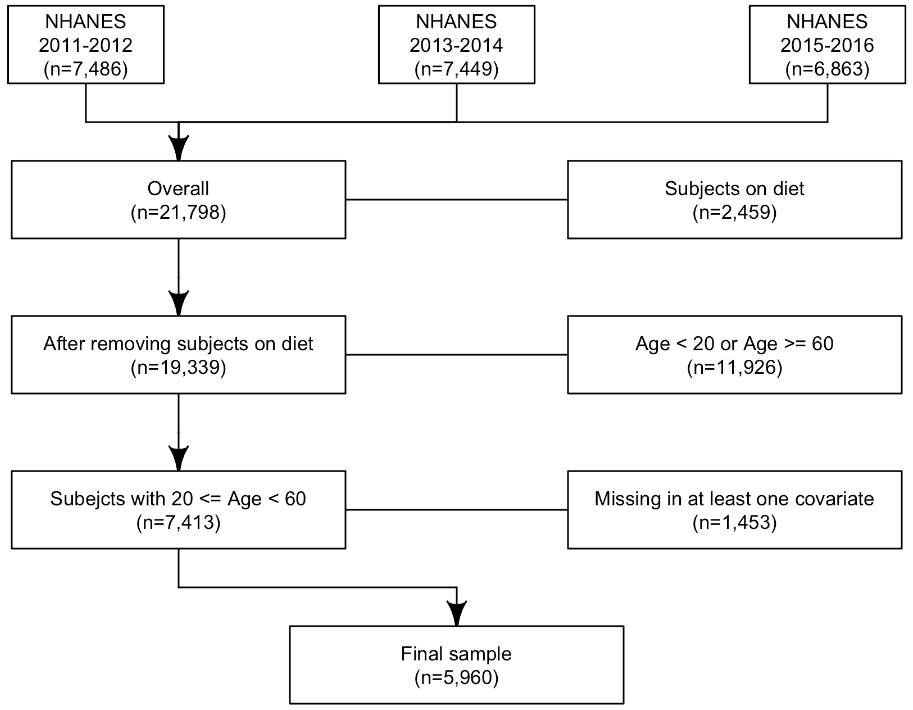

FIGURE 2
Analytic sample selection's flowchart: flowchart of analytic sample selection in NHANES 2011–2016.

The models were adjusted by covariates including age, sex, race, education as the highest grade or level of school completed or the highest degree received (defined as a continuous variable score ranging from 1 = Less than 9th grade, to 5 = College graduate or above), the ratio of family income to poverty (using the Department of Health and Human Services poverty guidelines), the minutes of sedentary activity represented by the time spent sitting on a typical day and the minutes of moderate (defined as activity that causes small increases in breathing or heart rate and is done for at least 10 min continuously) and vigorous activities (defined as activity that causes large increases in breathing or heart rate for at least 10 min continuously) spent either during work or during recreational activities on a typical day categorized using its tertiles (because of the skewed distribution), the smoking status as never-smokers (subjects who did not smoke as many as 100 cigarettes in their lifetime), former smokers (those who smoked at least 100 cigarettes in their lifetime but were not currently smoking cigarettes), and current smokers (subjects that currently smoked cigarettes) and the study cycle.

The Kruskal-Wallis test and Chi-squared test were used to test differences between obese and non-obese participants for continuous and categorical variables, respectively. A 2iWQS regression with repeated holdout was fit to assess the effect of the nutrients on obesity. We set 100 different training and validation splits of the dataset and 100 bootstrap iterations performed on each training set. All simulation and case study analyses can be replicated through the code in the Supplementary material.

Results

Simulation study

As a first step we tested for the best shrinkage parameter λ: a WQS regression was fitted on each dataset letting λ vary among the values 0, 1, 10, …, 10^4. The best shrinkage parameter was selected which minimized the AIC. In our case, $\lambda = 100$ was the optimum (i.e., smallest) value as shown in Figure 3.

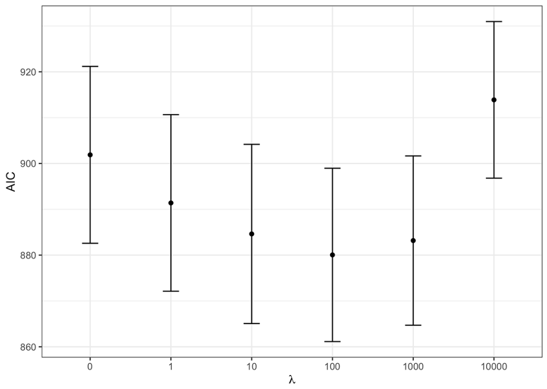

FIGURE 3
AIC at varying shrinkage parameters: WQS regression Akaike Information Criterion (AIC) at varying shrinkage parameter λ. The dots represent the average AIC obtained by fitting WQS models on the 100 datasets fixing λ to the corresponding value. Error bars are estimated as average AIC±one standard error.

We then checked the accuracy of the parameter estimates at varying shrinkage values. In Figure 4 the bias of the regression parameters β_{1p} and β_{1n} is presented for the positive and the negative direction, respectively, at varying λ. The most accurate estimates for the regression parameters correspond to the shrinkage parameter that minimized the AIC ($\lambda = 100$).

We performed the same analysis for the weights: in Figure 5 we can see that when $\lambda = 100$ we have accurate estimates both in terms of average sensitivity and specificity in identifying the elements truly associated with the outcome and those that do not have any relationship both for the positive (panel A) and negative (panel B) direction (positive direction: average sensitivity = 80.0%, average specificity = 92.9%; negative direction: average sensitivity = 77.2%, average specificity = 89.6%).

Once the optimum shrinkage parameter λ was identified, three additional regressions were fitted on each of the 100 datasets: in method 1 two separate WQS regressions without penalization were performed, one exploring the positive direction (i.e., where β_1 was constrained to be positive in the nonlinear estimation) and the second exploring the negative direction; in method 2 the WQS set of weights was estimated separately and without penalization for the positive and negative directions and once the WQS indices were estimated they were included in the same regression model fitted on the validation set; while in method 3 we applied quantile g-computation. We then compared these results with the ones obtained by applying the 2iWQS regression and penalized weights which we refer to as method 4.

Figure 6 shows the box-plots of the bias associated with the β_{1p} and β_{1n} for method 1, 2 and 4 and Ψ_{1p} and Ψ_{1n} for method 3. We can see how method 4 is able to give a better estimate of the parameters [Positive direction Median Error (PME): 0.022, Standard error (SE): 0.029; Negative direction Median Error (NME): −0.044, SE: 0.031] compared to method 1 (PME: −0.227, SE: 0.028; NME: 0.215, SE: 0.037), method 2 (PME: −0.109, SE: 0.025; NME: 0.065, SE: 0.037) and method 3 (PME: 0.318, SE: 0.080; NME: 0.310, SE: 0.080).

To measure the performance of the 3 methods in identifying the elements associated with the outcome and those that do not have any relationship (in this case method 1 and 2 share the same weight estimates) we considered the average sensitivity and specificity. When we set a threshold equal to the inverse of the number of elements in the mixture to discriminate between the significant and non-significant weights, we observed that method 4 shows better sensitivity in both directions compared to methods 1 and 2 (positive direction: method 1–2 = 76.0%, method 4 = 80.0%; negative direction: method 1–2 = 59.0%, method 4 = 77.2%) while the specificity was similar in both directions (positive direction: method 1–2 = 92.5%, method 4 = 92.9%; negative direction: method 1–2 = 90.2%, method 4 = 89.6%) (Supplementary Figures S1, S2). Method 3 was not considered in these comparisons since in quantile g-computation a rule to identify the significant elements in the association is not defined.

The same analysis was performed in a different scenario where the correlation values among the elements in the mixture were halved. Similar results were obtained in this scenario compared to the one considering the original correlation matrix: best estimates were observed in estimating beta regression parameters in the new scenario

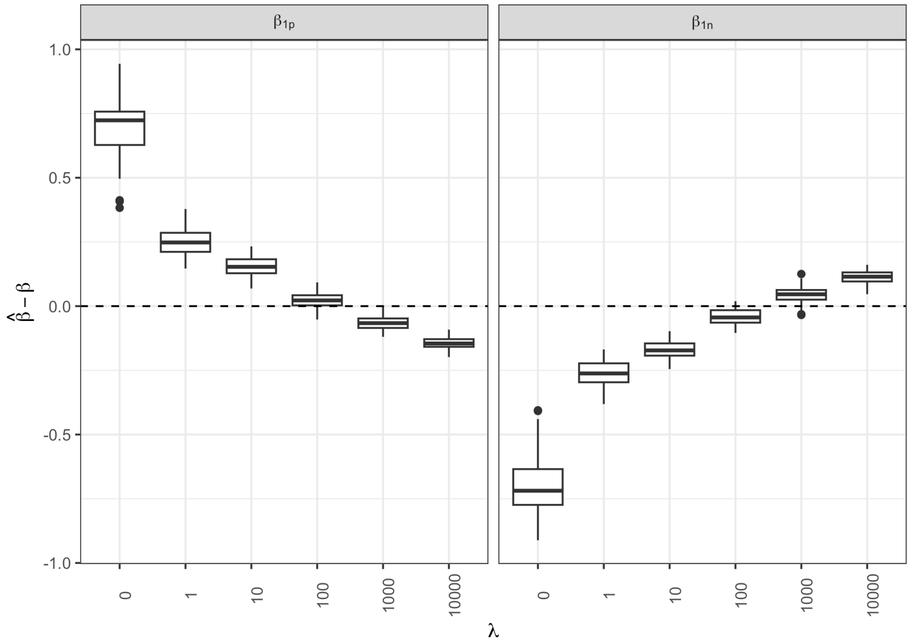

FIGURE 4
Regression parameter estimates at varying shrinkage parameters: box-plots of the bias associated to the β_{1p} and β_{1n} regression parameter estimates for the positive and negative direction, respectively, at different shrinkage parameters λ.

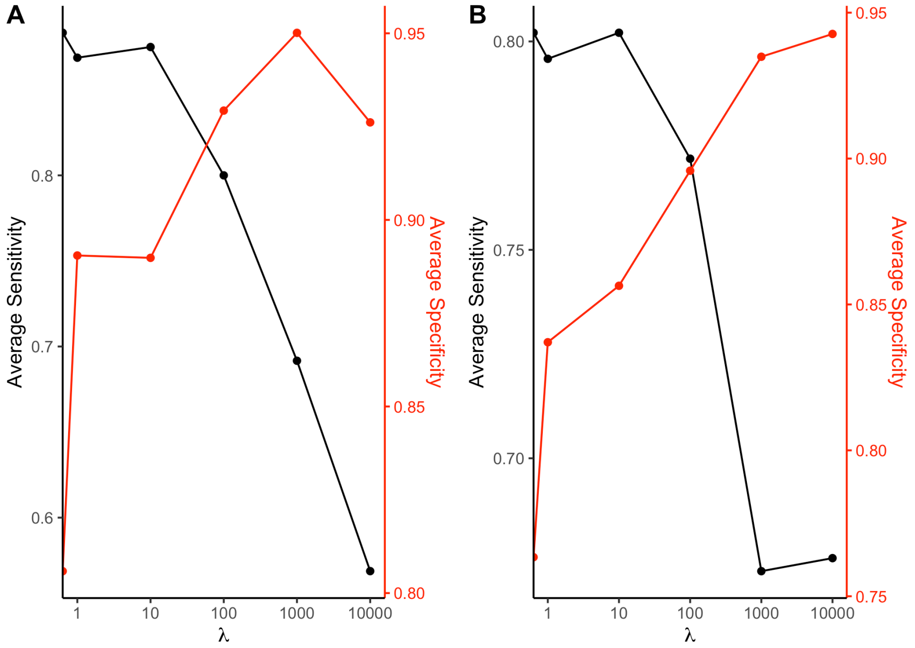

FIGURE 5
Sensitivity and specificity at varying shrinkage parameter: average sensitivity and specificity of the 2iWQS method in detecting the elements with a weight greater than 0 and those with null weight associated to a positive **(A)** or a negative **(B)** direction at varying λ.

for method 4 (PME: −0.018, SE: 0.042; NME: 0.028, SE: 0.052) and method 2 (PME: 0.001, SE: 0.035; NME: −0.037, SE:0.038), while method 1 (PME: −0.154, SE: 0.031; NME: 0.215, SE: 0.045) and method 3 (PME: 0.182, SE: 0.044; NME: −0.175, SE: 0.043) showed higher bias (Supplementary Figure S3).

When we looked at the ability of the methods in detecting the true elements with a non-null weight and those not associated with the outcome we observed a similar average sensitivity of method 4 and method 1–2 in both directions (positive direction: method 1–2 = 79.8%, method 4 = 80.2%; negative direction: method

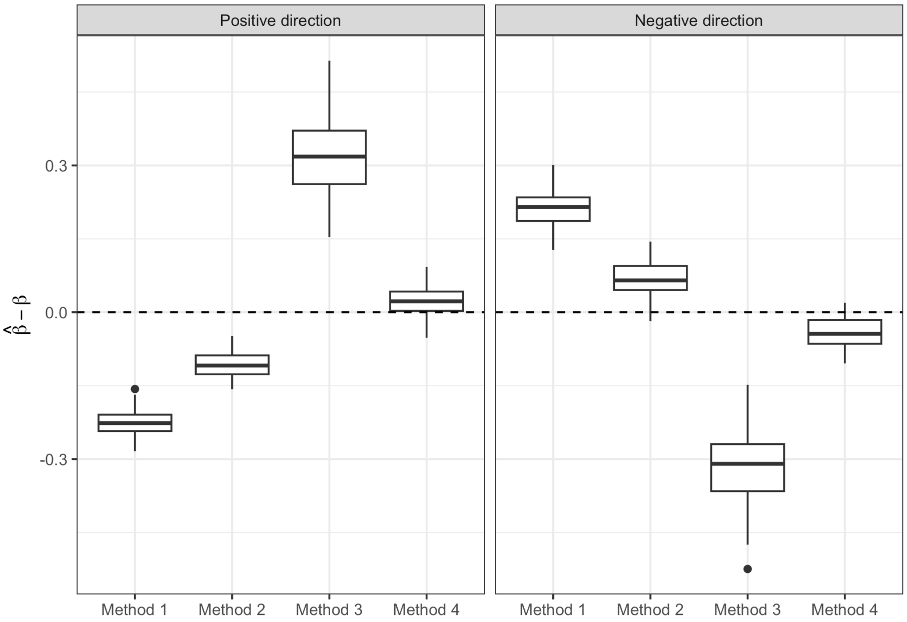

FIGURE 6
Regression parameter estimates of the three methods: box-plots of the bias in the regression parameter estimates associated with the two WQS indices of method 1, 2, and 4 and to the positive and negative ψ of quantile g-computation (method 3).

1–2 = 76.5%, method 4 = 78.7%) but better specificity of method 4 compared to method 1–2 in both directions (positive direction: method 1–2 = 94.3%, method 4 = 97.0%; negative direction: method 1–2 = 88.8%, method 4 = 95.5%) (Supplementary Figures S4, S5).

In the last analysis, we tested the performance of the penalized weights when considering a unidirectional association between the mixture and the outcome. In this case, we applied an additional method that included a single index exploring the positive direction (method 4 1d) and a penalization term was fixed at $\lambda = 1000$. We note that in the 2iWQS regression model if no set of weights is found in one of the two directions, we let the function automatically compute the single index WQS regression. A warning message saying that no weight estimates are found in one of the two directions is shown by the function. Method 4 showed better estimates in both directions of the regression parameter associated with the WQS index (method 4, PME: −0.033, SE: 0.024; NME: −0.026, SE: 0.026) compared to method 1 (PME: 0.052, SE: 0.026; NME: 0.151, SE: 0.062), method 2 (PME: 0.097, SE: 0.031; NME: −0.094, SE: 0.023) and method 3 (PME: 0.420, SE: 0.083; NME: −0.414, SE: 0.078) (Supplementary Figure S6). To evaluate the weight estimates we only considered the positive direction: Method 4 showed lower sensitivity (method 1–2 = 84.6%, method 4 = 60.8%) (Supplementary Figure S7) but similar specificity compared to methods 1–2 (method 1–2 = 92.1%, method 4 = 92.7%) (Supplementary Figure S8). Since no association was found in the negative direction (only in 1% of the scenarios we observed a false positive through method 4) we applied method 4 considering a single index in the positive direction (method 4 1d): a better estimate of the regression parameter was found (ME: −0.009, SE: 0.021) (Supplementary Figure S6) as well as a higher sensitivity (78.5%) (Supplementary Figure S7) and specificity (96.1%) (Supplementary Figure S8) compared to method 4, suggesting that if the association between the mixture and the dependent variable was not significant in one of the two directions, then a model with single index (the one exploring the significant direction) should be fitted as a final model to have more accurate estimates.

Case study

In total 5,960 subjects were included in the case study analysis. Table 2 shows the descriptive statistics of the overall population and divided by obese and non-obese for the covariates included in the 2iWQS regression. A total of 2,158 (36.2%) subjects were obese and were characterized by a higher median age and higher prevalence of females compared to non-obese participants. A different race distribution was also observed showing a higher percentage of Mexican and Black subjects and a lower prevalence of Asian participants among obese. Finally, a lower level of education was detected among obese people as well as a lower income to poverty ratio index, a higher number of minutes spent in sedentary activities and a higher frequency of low time spent performing moderate to vigorous-intensity activities. Summary statistics related to the 38 nutrients included in the analysis are shown in Table 3. All the elements that showed a significant difference between the two groups had higher values among non-obese subjects apart from caffeine.

We then applied the 2iWQS regression to test for the association between the nutrients and the outcome adjusting for all the covariates reported in Table 2. We used a repeated holdout approach to have more stable results including all the observations in the study to estimate both the weights and the regression parameters during the repeated testing and validation steps (25). A total of 100 repeated holdout 2iWQS regressions were performed. Table 4 shows the effects and their 95% Confidence Intervals (CIs) of both the positive and the negative index on the probability of being obese. Both indices were associated with the outcome. The median was used as the parameter point estimates while the 2.5th and the 97.5th percentiles were considered to build the 95% CIs. In Table 4 are also shown the medians of the weights greater than the prespecified cutoff (1/38 = 0.026) for the positive and negative index where sodium and magnesium showed a predominant role in the positive and negative association with obesity, respectively. However, there is also an indication of a mixture effect in the positive and in the negative direction including other components with medians above the cutoff threshold. The distributions of all the elements included in the analysis estimated for both indices are provided in Figure 7.

Discussion

Through this work we were able to extend WQS regression to the case where the mixture considered in the study can have both a positive and a negative effect, moreover, we increased the ability of WQS in detecting the association between the mixture and the outcome as well as in identifying the true elements through the introduction of penalized weight estimates. The new method estimates two indices in the same regression model both including all the elements of the mixture, one constrained to be positive and one to be negative. A recent work from Keil and others (24) introduced a new approach that estimates the overall effect of the mixture on the outcome when there is uncertainty about the effect direction of some exposures. Differently from the original WQS regression, they proposed to estimate positive and negative weights within the same index using normalized linear (or generalized linear) regression coefficients and then estimating the effect of the overall mixture via a standard g-computation algorithm. With the q g-comp approach, the overall effect of the mixture is estimated (which is not the mixture effect) on the outcome, but we cannot estimate the impact in the positive and negative directions, especially with highly correlated components. Through our method, we introduced the possibility to measure both the beneficial and harmful effects of the exposure to the mixture separately keeping the advantages of the identification of a weighted index that allows increasing the power to detect the mixture effect and avoiding incurring the reversal paradox. This ability of the 2iWQS is at the expense of a small bias in the estimate of the regression parameter; however, we were able to almost null the bias due to the

TABLE 2 Sociodemographics and lifestyles: descriptive statistics of the variables included in the study for the overall population and divided by obese and non-obese.

	Non-obese (*N*=3,802)	Obese (*N*=2,158)	Total (*N*=5,960)	*p* value
Median (Q1, Q3)				
Age	38.0 (28.0, 48.0)	41.0 (32.0, 50.0)	39.0 (30.0, 49.0)	**<0.001**
Education	4.0 (3.0, 5.0)	4.0 (3.0, 4.0)	4.0 (3.0, 5.0)	**<0.001**
Family income to poverty ratio	2.4 (1.1, 4.6)	1.9 (1.0, 3.7)	2.2 (1.1, 4.2)	**<0.001**
Minutes of sedentary activity	360.0 (240.0, 480.0)	360.0 (240.0, 540.0)	360.0 (240.0, 480.0)	**<0.001**
Count (%)				
Sex				**<0.001**
Males	2003 (52.7%)	961 (44.5%)	2,964 (49.7%)	
Females	1799 (47.3%)	1,197 (55.5%)	2,996 (50.3%)	
Race				**<0.001**
Asian	703 (18.5%)	84 (3.9%)	787 (13.2%)	
Black	671 (17.6%)	564 (26.1%)	1,235 (20.7%)	
Mexican	428 (11.3%)	374 (17.3%)	802 (13.5%)	
Other Hispanic	351 (9.2%)	209 (9.7%)	560 (9.4%)	
Others	140 (3.7%)	87 (4.0%)	227 (3.8%)	
White	1,509 (39.7%)	840 (38.9%)	2,349 (39.4%)	
Moderate and vigorous-intensity activities				**<0.001**
Low	1,251 (32.9%)	887 (41.1%)	2,138 (35.9%)	
Medium	1,312 (34.5%)	592 (27.4%)	1,904 (31.9%)	
High	1,239 (32.6%)	679 (31.5%)	1,918 (32.2%)	
Smoking status				0.399
Never	2,320 (61.0%)	1,283 (59.5%)	3,603 (60.5%)	
Former	615 (16.2%)	375 (17.4%)	990 (16.6%)	
Current	867 (22.8%)	500 (23.2%)	1,367 (22.9%)	

Median, 1st (Q1) and 3rd quartiles (Q3) are shown for continuous variables while counts and percentages were considered for categorical variables. Kruskal-Wallis test and Chi-squared test were used to test for differences for continuous and categorical variables, respectively. Bolded *p*-values highlight the statistically significant results.

TABLE 3 Nutrient characteristics: summary statistics of the 38 nutrients included in the analysis.

	Non-obese (*N*=3,802)	Obese (*N*=2,158)	Total (*N*=5,960)	*p* value
Added vitamin B12 (mcg)	0.0 (0.0, 1.2)	0.0 (0.0, 0.9)	0.0 (0.0, 1.1)	**0.005**
Alcohol (gm)	0.0 (0.0, 7.8)	0.0 (0.0, 0.0)	0.0 (0.0, 5.6)	**<0.001**
Alpha-carotene (mcg)	70.0 (20.0, 443.4)	51.0 (15.0, 248.0)	62.2 (18.0, 370.2)	**<0.001**
Beta-carotene (mcg)	1091.5 (417.6, 2998.8)	798.0 (338.9, 2041.6)	975.0 (382.0, 2565.0)	**<0.001**
Caffeine (mg)	94.0 (25.5, 190.0)	100.2 (29.0, 204.4)	96.0 (27.5, 193.0)	**0.022**
Calcium (mg)	958.8 (656.1, 1358.4)	941.0 (639.6, 1316.6)	952.1 (650.9, 1347.5)	0.120
Carbohydrate (gm)	254.7 (190.2, 332.9)	245.0 (183.9, 318.8)	250.2 (188.3, 327.3)	**0.001**
Cholesterol (mg)	254.5 (158.0, 393.9)	262.0 (164.0, 410.0)	256.5 (160.0, 400.0)	0.050
Total choline (mg)	314.1 (224.0, 434.8)	302.4 (213.2, 423.4)	310.6 (220.0, 430.5)	**0.002**
Copper (mg)	1.3 (0.9, 1.8)	1.1 (0.8, 1.6)	1.2 (0.9, 1.8)	**<0.001**
Beta-cryptoxanthin (mcg)	38.5 (12.5, 101.0)	37.0 (12.5, 93.5)	37.5 (12.5, 98.5)	0.258
Dietary fiber (gm)	16.4 (11.2, 23.9)	14.6 (10.1, 21.0)	15.8 (10.7, 22.8)	**<0.001**
Folate, DFE (mcg)	606.8 (403.5, 949.7)	527.0 (345.1, 829.1)	578.0 (381.0, 908.8)	**<0.001**
Iron (mg)	15.1 (10.9, 21.7)	14.0 (10.0, 20.5)	14.7 (10.5, 21.2)	**<0.001**
Lutein + zeaxanthin (mcg)	891.2 (474.5, 1824.0)	785.2 (424.6, 1471.2)	854.8 (455.0, 1681.2)	**<0.001**
Lycopene (mcg)	2536.5 (644.0, 6611.5)	2498.8 (600.0, 6545.4)	2525.5 (622.8, 6598.8)	0.564
Magnesium (mg)	307.0 (229.5, 414.0)	281.0 (202.8, 370.0)	297.4 (219.0, 397.6)	**<0.001**
Total monounsaturated fatty acids (gm)	26.9 (18.7, 36.5)	26.8 (19.2, 36.2)	26.9 (18.9, 36.4)	0.844
Niacin (mg)	28.5 (20.2, 40.4)	26.6 (18.8, 37.7)	27.8 (19.7, 39.4)	**<0.001**
Phosphorus (mg)	1352.2 (1022.6, 1763.0)	1295.5 (986.6, 1723.4)	1332.8 (1009.0, 1745.5)	**0.003**
Total polyunsaturated fatty acids (gm)	17.4 (11.9, 24.3)	17.6 (12.2, 24.8)	17.4 (12.0, 24.5)	0.375
Potassium (mg)	2598.2 (1954.2, 3360.9)	2399.5 (1818.2, 3086.4)	2526.2 (1898.6, 3281.9)	**<0.001**
Protein (gm)	81.7 (61.3, 107.5)	78.0 (58.3, 102.5)	80.5 (60.1, 105.7)	**<0.001**
Riboflavin (Vitamin B2) (mg)	2.2 (1.5, 3.3)	2.1 (1.4, 3.1)	2.1 (1.5, 3.3)	**<0.001**
Total saturated fatty acids (gm)	24.0 (16.2, 34.1)	24.4 (17.1, 34.3)	24.2 (16.4, 34.2)	0.081
Selenium (mcg)	122.8 (89.0, 166.4)	116.2 (84.2, 159.8)	120.6 (87.1, 163.9)	**<0.001**
Sodium (mg)	3477.2 (2582.1, 4578.4)	3401.2 (2585.0, 4410.8)	3456.8 (2582.9, 4522.5)	0.110
Total sugars (gm)	100.4 (65.6, 145.2)	102.9 (66.7, 148.1)	101.1 (66.0, 146.4)	0.277
Theobromine (mg)	7.5 (0.0, 41.5)	7.0 (0.0, 40.5)	7.5 (0.0, 41.0)	0.184
Thiamin (Vitamin B1) (mg)	1.8 (1.3, 2.7)	1.6 (1.1, 2.4)	1.7 (1.2, 2.6)	**<0.001**
Vitamin A, RAE (mcg)	517.0 (304.5, 816.0)	456.8 (275.1, 724.2)	495.0 (293.4, 774.0)	**<0.001**
Vitamin B12 (mcg)	5.5 (3.1, 10.6)	5.1 (2.9, 9.7)	5.3 (3.0, 10.3)	**0.003**
Vitamin B6 (mg)	2.4 (1.6, 3.9)	2.1 (1.4, 3.4)	2.3 (1.5, 3.7)	**<0.001**
Vitamin C (mg)	86.6 (37.5, 169.6)	70.5 (28.5, 143.1)	81.3 (33.9, 158.1)	**<0.001**
Vitamin D (D2 + D3) (mcg)	5.6 (2.3, 13.7)	4.6 (2.1, 11.2)	5.2 (2.2, 12.9)	**<0.001**
Vitamin E as alpha-tocopherol (mg)	7.7 (5.3, 11.3)	7.1 (4.9, 10.4)	7.5 (5.2, 10.9)	**<0.001**
Vitamin K (mcg)	87.0 (51.4, 154.0)	74.8 (44.5, 127.9)	82.7 (48.9, 143.3)	**<0.001**
Zinc (mg)	12.3 (8.4, 18.3)	11.6 (7.8, 17.1)	12.0 (8.2, 17.7)	**<0.001**

Median, 1st and 3rd quartiles are shown for the overall population and divided in obese and non-obese participants. Kruskal-Wallis test was applied to test for differences between the two groups. Bolded *p*-values highlight the statistically significant results.

inclusion of both indices in the same model and the penalized estimates of the weights. Another advantage that we carry from WQS regression compared to other shrinkage methods is the ability to estimate a mixture effect, in the 2iWQS case in both directions, and with highly correlated variables to avoid a grouping effect like in elastic net or the arbitrary selection of variables like in LASSO that can be problematic in the risk evaluation of environmental mixture (19).

In this work, we showed how the two indices were built and how we deal with the correlation between the two indices. This was the main issue that we encountered: because of the high correlation among the elements included in the mixture we noticed a risk of collinearity when including both indices in the same regression model. To address this problem, we applied two strategies. As a first solution, we introduced a penalization parameter in the objective

TABLE 4 2iWQS regression results: the estimates and 95% confidence intervals (CI) of the positive (PWQS) and negative (NWQS) indices are shown. The second part of the table shows the magnitude of the weights greater than the prespecified cutoff (0.026) for both the positive and the negative index.

	Estimate	95% CI
PWQS	0.107	0.051; 0.166
NWQS	−0.156	−0.194; −0.101
	Weights	
PWQS		
Sodium	0.199	
Polyunsaturated fat	0.077	
Cholesterol	0.059	
Caffeine	0.038	
Calcium	0.028	
NWQS		
Magnesium	0.172	
Fiber	0.081	
Vitamin C	0.075	
Vitamin B6	0.059	
Beta-carotene	0.049	
Folate DFE	0.048	
Vitamin D	0.046	
Vitamin E	0.040	
Vitamin K	0.036	
Alpha-carotene	0.034	

The model was adjusted for age, sex, race, education, the ratio of family income to poverty, the minutes of sedentary activity, the minutes of moderate and vigorous activities and the smoking status. CI, Confidence Interval; DFE, Dietary Folate Equivalents; PWQS, Positive Weighted Quantile Sum index; NWQS, Negative Weighted Quantile Sum index.

function to estimate the weights allowing us to better discriminate between the elements that have a weight significantly different from zero and those that have a null weight. This reduces the noise produced by the elements that are not associated with the outcome which can increase the correlation between the two indices. The second solution to reduce the correlation between the two indices was the application of the weighted mean based on the tolerance of each bootstrapped model in the estimates of the final weights. This gave more importance to the set of weights that produces less correlated indices.

In the real case study, we applied this new methodology by taking data from the NHANES 2011–2012, 2013–2014, and 2015–2016 study cycles to test for the association between nutrients and obesity. The results showed a harmful effect of sodium, polyunsaturated fatty acids, cholesterol, caffeine and calcium. While some nutrients like sodium (29–31) and cholesterol (32, 33) are already known risk elements for obesity, we also found nutrients for which there is controversial evidence of their effect on obesity. Because of the unavailable information on the different types of polyunsaturated fatty acids we were not able to disentangle which component drove the harmful effect on obesity. Polyunsaturated fatty acids are known to be protective against overweight and obesity (34–39), however, there is evidence that an increased intake of omega-6 long-chain polyunsaturated fatty acids can increase the risk of obesity (35, 40) in particular if there is an unbalanced omega-6/omega-3 ratio, an increasingly widespread problem in western countries (41). Calcium intake was observed to be a protective factor against obesity (42–45) but few studies did not show any effect (46, 47) or a harmful relationship (48). Finally, caffeine has a protective effect on a regular intake (49) but in excessive doses, it can affect insomnia and anxiety (50, 51) which are associated in turn with obesity (52–54). On the other hand, a protective effect against obesity was found for magnesium, fiber, vitamin C, vitamin B6, beta-carotene, folate Dietary Folate Equivalents (DFE), vitamin D, vitamin E, vitamin K and

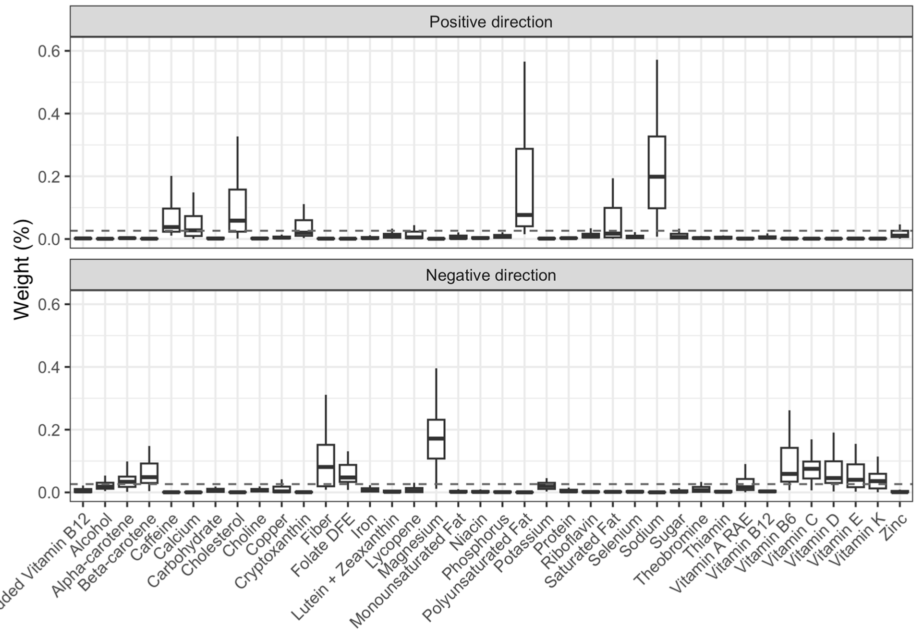

FIGURE 7
Box-plot of the weights associated to the positive and negative index estimate through the repeated holdout 2iWQS regression. The dashed line represents the prespecified cut-off established to identify the most important elements of the mixture and set equal to the inverse of the number of the mixture components (0.026).

alpha-carotene. For these nutrients there was evidence of a beneficial effect against obesity in previous studies (55–70).

One strength of this novel approach is the ability to include all nutrients in the analysis considering the possible confounding that can be caused by the exclusion of some elements while previous studies showed the association of one or few elements at a time with obesity. Moreover, we showed that considering two indices in the same WQS regression with the penalization term increased the accuracy of the parameter estimates when the mixture has a bidirectional effect on the outcome of interest. In addition to already available methods like the quantile-based g-computation approach, our new methodology allows us to measure the double association of the mixture quantifying the effect in both the positive and negative direction with the dependent variable. As for WQS regression, a further advantage of this approach is the ease of use and of interpretation of the results due to the building of the indices representing the mixture exposure and the weights attributed to each element identifying the contribution of each element to the estimated association with the outcome. One additional advantage to be considered is the possibility to integrate this new feature of WQS in the random subset (21) and repeated holdout (25) extensions.

In contrast to the simplicity of the method, we can identify as a limitation of the WQS regression the lower flexibility due to the assumption of a linear trend between each element and the dependent variable. Before applying this method, it is recommended to check in advance if a non-linear trend exists between the mixture components and the outcome by adding a quadratic term in the WQS model to test for curvilinearity or by applying other regression methods that can deal with environmental mixtures like Bayesian Kernel Machine Regression (71).

Conclusion

Through this work, we introduced an extension of WQS regression that improved the accuracy of the parameter estimates in WQS with a single index as well as when considering a mixture of elements where a subset may have a protective effect and other components may have a harmful effect on the outcome. This was possible through the application of penalized weight estimates and the introduction of a second index, allowing one WQS index to investigate the positive and one the negative direction of the association with the dependent variable. The usage of two indices in the same model quantifies the effect of the mixture in the two directions keeping them separate. This can be of interest in fields like nutrition where a measure of the protective and harmful effects of nutrients can help in dietary decision making.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://cran.r-project.org/web/packages/gWQS/index.html.

Ethics statement

The studies involving human participants were reviewed and approved by NCHS Research Ethics Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

SR gave substantial contributions to the conception, the analysis and interpretation of data for the work, drafted the work and revised it critically for important intellectual content, and provided approval for publication of the content. CG and SC gave substantial contributions to the conception and interpretation of data for the work, revised the work critically for important intellectual content, and provided approval for publication of the content. All authors contributed to the article and approved the submitted version.

Funding

SR was supported by Programma Operativo Nazionale "Ricerca e Innovazione" 2014–2020 (PON R&I FSE-REACT EU), Azione IV.4 "Contratti di ricerca su tematiche dell'innovazione." CG was supported by the National Institute of Environmental Health Sciences (#U2CES026555, #P30ES023515, and #R01ES028811). SC was supported by research grants from the Italian Ministry of University (PRIN projects n. 20178S4EK9).

Acknowledgments

The authors would like to acknowledge support from the Italian Ministry of University and Research through its program Programma Operativo Nazionale "Ricerca e Innovazione" 2014–2020 (PON R&I FSE-REACT EU), Azione IV.4 "Contratti di ricerca su tematiche dell'innovazione" and the project PRIN projects n. 20178S4EK9 and the National Institute of Environmental Health Sciences (grants numbers #U2CES026555 and #R01ES028811).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2023.1151821/full#supplementary-material

References

1. Carpenter DO, Arcaro K, Spink DC. Understanding the human health effects of chemical mixtures. *Environ Health Perspect*. (2002) 110:25–42. doi: 10.1289/ehp.02110s125

2. Carlin DJ, Rider CV, Woychik R, Birnbaum LS. Unraveling the health effects of environmental mixtures: an NIEHS priority. *Environ Health Perspect*. (2013) 121:A6–8. doi: 10.1289/ehp.1206182

3. Patel CJ. Analytic complexity and challenges in identifying mixtures of exposures associated with phenotypes in the exposome era. *Curr Epidemiol Rep*. (2017) 4:22–30. doi: 10.1007/s40471-017-0100-5

4. Stafoggia M, Breitner S, Hampel R, Basagaña X. Statistical approaches to address multi-pollutant mixtures and multiple exposures: the state of the science. *Curr Environ Health Rep*. (2017) 4:481–90. doi: 10.1007/s40572-017-0162-z

5. Braun JM, Gennings C, Hauser R, Webster TF. What can epidemiological studies tell us about the impact of chemical mixtures on human health? *Environ Health Perspect*. (2016) 124:A6–9. doi: 10.1289/ehp.1510569

6. Greenland S. Multiple comparisons and association selection in general epidemiology. *Int J Epidemiol*. (2008) 37:430–4. doi: 10.1093/ije/dyn064

7. Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*. (2005) 61:738–48. doi: 10.1111/j.1541-0420.2005.00344.x

8. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. *Am J Epidemiol*. (1995) 142:904–8. doi: 10.1093/oxfordjournals.aje.a117737

9. Stacey AW, Czyz CN. Author response: analysis of the use of multiple comparison corrections in ophthalmology research. *Invest Ophthalmol Vis Sci*. (2012) 53:5955. doi: 10.1167/iovs.12-10642

10. Dominici F, Peng RD, Barr CD, Bell ML. Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. *Epidemiology*. (2010) 21:187–94. doi: 10.1097/EDE.0b013e3181cc86e8

11. Dormann CF, Elith J, Bacher S, Carré GCG, García Márquez JR, Gruber B, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. (2013) 36:27–46. doi: 10.1111/j.1600-0587.2012.07348.x

12. Vatcheva KP, Lee M, McCormick JB, Rahbar MH. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology*. (2016) 6:227. doi: 10.4172/2161-1165.1000227

13. Leal C, Bean K, Thomas F, Chaix B. Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable. *Am J Epidemiol*. (2012) 175:1152–62. doi: 10.1093/aje/kwr434

14. Weisskopf MG, Seals RM, Webster TF. Bias amplification in epidemiologic analysis of exposure to mixtures. *Environ Health Perspect*. (2018) 126:047003. doi: 10.1289/EHP2450

15. Jain P, Vineis P, Liquet B, Vlaanderen J, Bodinier B, van Veldhoven K, et al. A multivariate approach to investigate the combined biological effects of multiple exposures. *J Epidemiol Community Health*. (2018) 72:564–71. doi: 10.1136/jech-2017-210061

16. Sun Z, Tao Y, Li S, Ferguson KK, Meeker JD, Park SK, et al. Statistical strategies for constructing health risk models with multiple pollutants and their interactions: possible choices and comparisons. *Environ Health*. (2013) 12:85. doi: 10.1186/1476-069X-12-85

17. Billionnet C, Sherrill D, Annesi-Maesano I, Study G. Estimating the health effects of exposure to multi-pollutant mixture. *Ann Epidemiol*. (2012) 22:126–41. doi: 10.1016/j.annepidem.2011.11.004

18. Tu Y, Gunnell D. Gilthorpe MS. Simpson's paradox, Lord's paradox, and suppression effects are the same phenomenon – the reversal paradox. *Emerg Themes Epidemiol*. (2008) 5:2. doi: 10.1186/1742-7622-5-2

19. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. *J Agric Biol Environ Stat*. (2015) 20:100–20. doi: 10.1007/s13253-014-0180-3

20. Czarnota J, Gennings C, Wheeler DC. Assessment of weighted quantile sum regression for modeling chemical mixtures and cancer risk. *Cancer Inform*. (2015) 14:159–71. doi: 10.4137/CIN.S17295

21. Curtin P, Kellogg J, Cech N, Gennings C. A random subset implementation of weighted Quantile sum (WQSRS) regression for analysis of high-dimensional mixtures. *Commun Stat Simulat Comput*. (2019) 50:1119–34. doi: 10.1080/03610918.2019.1577971

22. Eggers S, Bixby M, Renzetti S, Curtin P, Gennings C. Human microbiome mixture analysis using weighted Quantile sum regression. *Int J Environ Res Public Health*. (2022) 20:94. doi: 10.3390/ijerph20010094

23. Renzetti S, Curtin P, Just AC, Bello G, Gennings C. *gWQS: generalized weighted Quantile sum regression*. CRAN repository (2021).

24. Keil AP, Buckley JP, O'Brien KM, Ferguson KK, Zhao S, White AJ. A Quantile-based g-computation approach to addressing the effects of exposure mixtures. *Environ Health Perspect*. (2020) 128:47004. doi: 10.1289/EHP5838

25. Tanner EM, Bornehag CG, Gennings C. Repeated holdout validation for weighted quantile sum regression. *MethodsX*. (2019) 6:2855–60. doi: 10.1016/j.mex.2019.11.008

26. Day DB, Sathyanarayana S, LeWinn KZ, Karr CJ, Mason WA, Szpiro AA. A permutation test-based approach to strengthening inference on the effects of environmental mixtures: comparison between single-index analytic methods. *Environ Health Perspect*. (2022) 130:87010. doi: 10.1289/EHP10570

27. National Health and Nutrition Examination Survey (NHANES). *MEC in-person dietary interviewers procedures manual* Centers for Disease Control and Prevention (CDC) (2016).

28. National Health and Nutrition Examination Survey (NHANES). *Phone follow-up dietary interviewer procedures manual*. Centers for Disease Control and Prevention (CDC) (2016).

29. Kang YJ, Wang HW, Cheon SY, Lee HJ, Hwang KM, Yoon HS. Associations of obesity and dyslipidemia with intake of sodium, fat, and sugar among Koreans: a qualitative systematic review. *Clin Nutr Res*. (2016) 5:290–304. doi: 10.7762/cnr.2016.5.4.290

30. Zhou L, Stamler J, Chan Q, Van Horn L, Daviglus ML, Dyer AR, et al. Salt intake and prevalence of overweight/obesity in Japan, China, the United Kingdom, and the United States: the INTERMAP Study. *Am J Clin Nutr*. (2019) 110:34–40. doi: 10.1093/ajcn/nqz067

31. Lee J, Hwang Y, Kim KN, Ahn C, Sung HK, Ko KP, et al. Associations of urinary sodium levels with overweight and central obesity in a population with a sodium intake. *BMC Nutr*. (2018) 4:47. doi: 10.1186/s40795-018-0255-6

32. Tall AR, Yvan-Charvet L. Cholesterol, inflammation and innate immunity. *Nat Rev Immunol*. (2015) 15:104–16. doi: 10.1038/nri3793

33. Sozen E, Ozer NK. Impact of high cholesterol and endoplasmic reticulum stress on metabolic diseases: an updated mini-review. *Redox Biol*. (2017) 12:456–61. doi: 10.1016/j.redox.2017.02.025

34. Tortosa-Caparrós E, Navas-Carrillo D, Marín F, Orenes-Piñero E. Anti-inflammatory effects of omega 3 and omega 6 polyunsaturated fatty acids in cardiovascular disease and metabolic syndrome. *Crit Rev Food Sci Nutr*. (2017) 57:3421–9. doi: 10.1080/10408398.2015.1126549

35. Saini RK, Keum YS. Omega-3 and omega-6 polyunsaturated fatty acids: dietary sources, metabolism, and significance – a review. *Life Sci*. (2018) 203:255–67. doi: 10.1016/j.lfs.2018.04.049

36. Ralston JC, Lyons CL, Kennedy EB, Kirwan AM, Roche HM. Fatty acids and NLRP3 inflammasome-mediated inflammation in metabolic tissues. *Annu Rev Nutr*. (2017) 37:77–102. doi: 10.1146/annurev-nutr-071816-064836

37. Rogero MM, Calder PC. Obesity, inflammation, toll-like receptor 4 and fatty acids. *Nutrients*. (2018) 10:432. doi: 10.3390/nu10040432

38. Silva Figueiredo P, Carla Inada A, Marcelino G, Maiara Lopes Cardozo C, de Cássia FK, de Cássia Avellaneda Guimarães R, et al. Fatty acids consumption: the role metabolic aspects involved in obesity and its associated disorders. *Nutrients*. (2017) 9:1158. doi: 10.3390/nu9101158

39. Albracht-Schulte K, Kalupahana NS, Ramalingam L, Wang S, Rahman SM, Robert-McComb J, et al. Omega-3 fatty acids in obesity and metabolic syndrome: a mechanistic update. *J Nutr Biochem*. (2018) 58:1–16. doi: 10.1016/j.jnutbio.2018.02.012

40. Fekete K, Györei E, Lohner S, Verduci E, Agostoni C, Decsi T. Long-chain polyunsaturated fatty acid status in obesity: a systematic review and meta-analysis. *Obes Rev*. (2015) 16:488–97. doi: 10.1111/obr.12280

41. Simopoulos AP. An increase in the Omega-6/Omega-3 fatty acid ratio increases the risk for obesity. *Nutrients*. (2016) 8:128. doi: 10.3390/nu8030128

42. Pannu PK, Calton EK, Soares MJ. Calcium and vitamin D in obesity and related chronic disease. *Adv Food Nutr Res*. (2016) 77:57–100. doi: 10.1016/bs.afnr.2015.11.001

43. Zhang F, Ye J, Zhu X, Wang L, Gao P, Shu G, et al. Anti-obesity effects of dietary calcium: the evidence and possible mechanisms. *Int J Mol Sci*. (2019) 20:3072. doi: 10.3390/ijms20123072

44. Villarroel P, Villalobos E, Reyes M, Cifuentes M. Calcium, obesity, and the role of the calcium-sensing receptor. *Nutr Rev*. (2014) 72:627–37. doi: 10.1111/nure.12135

45. de Oliveira Freitas DM, Stampini Duarte Martino H, Machado Rocha Ribeiro S, Gonçalves Alfenas RC. Calcium ingestion and obesity control. *Nutr Hosp*. (2012) 27:1758–71. doi: 10.3305/nh.2012.27.6.5977

46. Soares MJ, Chan She Ping-Delfos W, Ghanbari MH. Calcium and vitamin D for obesity: a review of randomized controlled trials. *Eur J Clin Nutr*. (2011) 65:994–1004. doi: 10.1038/ejcn.2011.106

47. Song Q, Sergeev IN. Calcium and vitamin D in obesity. *Nutr Res Rev*. (2012) 25:130–41. doi: 10.1017/S0954422412000029

48. Sadeghi O, Keshteli AH, Doostan F, Esmaillzadeh A, Adibi P. Association between dairy consumption, dietary calcium intake and general and abdominal obesity among Iranian adults. *Diabetes Metab Syndr*. (2018) 12:769–75. doi: 10.1016/j.dsx.2018.04.040

49. Bhatti SK, O'Keefe JH, Lavie CJ. Coffee and tea: perks for health and longevity? *Curr Opin Clin Nutr Metab Care*. (2013) 16:688–97. doi: 10.1097/MCO.0b013e328365b9a0

50. Nehlig A. Interindividual differences in caffeine metabolism and factors driving caffeine consumption. *Pharmacol Rev*. (2018) 70:384–411. doi: 10.1124/pr.117.014407

51. Yang A, Palmer AA, de Wit H. Genetics of caffeine consumption and responses to caffeine. *Psychopharmacology*. (2010) 211:245–57. doi: 10.1007/s00213-010-1900-1

52. Amiri S, Behnezhad S. Obesity and anxiety symptoms: a systematic review and meta-analysis. *Neuropsychiatrie*. (2019) 33:72–89. doi: 10.1007/s40211-019-0302-9

53. Rajan TM, Menon V. Psychiatric disorders and obesity: a review of association studies. *J Postgrad Med*. (2017) 63:182–90. doi: 10.4103/jpgm.JPGM_712_16

54. Cai GH, Theorell-Haglöw J, Janson C, Svartengren M, Elmståhl S, Lind L, et al. Insomnia symptoms and sleep duration and their combined effects in relation to associations with obesity and central obesity. *Sleep Med*. (2018) 46:81–7. doi: 10.1016/j.sleep.2018.03.009

55. Coronel J, Pinos I, Amengual J. β-Carotene in obesity research: technical considerations and current status of the field. *Nutrients*. (2019) 11:842. doi: 10.3390/nu11040842

56. Bonet ML, Canas JA, Ribot J, Palou A. Carotenoids in adipose tissue biology and obesity. *Subcell Biochem*. (2016) 79:377–414. doi: 10.1007/978-3-319-39126-7_15

57. Perveen R, Suleria HA, Anjum FM, Butt MS, Pasha I, Ahmad S. Tomato (*Solanum lycopersicum*) carotenoids and lycopenes chemistry; metabolism, absorption, nutrition, and allied health claims—a comprehensive review. *Crit Rev Food Sci Nutr*. (2015) 55:919–29. doi: 10.1080/10408398.2012.657809

58. Pereira-Santos M, Costa PR, Assis AM, Santos CA, Santos DB. Obesity and vitamin D deficiency: a systematic review and meta-analysis. *Obes Rev*. (2015) 16:341–9. doi: 10.1111/obr.12239

59. Savastano S, Barrea L, Savanelli MC, Nappi F, Di Somma C, Orio F, et al. Low vitamin D status and obesity: role of nutritionist. *Rev Endocr Metab Disord*. (2017) 18:215–25. doi: 10.1007/s11154-017-9410-7

60. Walsh JS, Bowles S, Evans AL. Vitamin D in obesity. *Curr Opin Endocrinol Diabetes Obes*. (2017) 24:389–94. doi: 10.1097/MED.0000000000000371

61. Pourshahidi LK. Vitamin D and obesity: current perspectives and future directions. *Proc Nutr Soc*. (2015) 74:115–24. doi: 10.1017/S0029665114001578

62. Garcia-Diaz DF, Lopez-Legarrea P, Quintero P, Martinez JA. Vitamin C in the treatment and/or prevention of obesity. *J Nutr Sci Vitaminol*. (2014) 60:367–79. doi: 10.3177/jnsv.60.367

63. Thomas-Valdés S, Tostes MDGV, Anunciação PC, da Silva BP, Sant'Ana HMP. Association between vitamin deficiency and metabolic disorders related to obesity. *Crit Rev Food Sci Nutr*. (2017) 57:3332–43. doi: 10.1080/10408398.2015.1117413

64. Şerban CL, Sima A, Hogea CM, Chiriţă-Emandi A, Perva IT, Vlad A, et al. Assessment of nutritional intakes in individuals with obesity under medical supervision. A cross-sectional Study. *Int J Environ Res Public Health*. (2019) 16:3036. doi: 10.3390/ijerph16173036

65. Oliveira AR, Cruz KJ, Severo JS, Morais JB, Freitas TE, Araújo RS, et al. Hypomagnesemia and its relation with chronic low-grade inflammation in obesity. *Rev Assoc Med Bras*. (2017) 63:156–63. doi: 10.1590/1806-9282.63.02.156

66. Piuri G, Zocchi M, Della Porta M, Ficara V, Manoni M, Zuccotti GV, et al. Magnesium in obesity, metabolic syndrome, and type 2 diabetes. *Nutrients*. (2021) 13:320. doi: 10.3390/nu13020320

67. Astrup A, Bügel S. Overfed but undernourished: recognizing nutritional inadequacies/deficiencies in patients with overweight or obesity. *Int J Obes*. (2019) 43:219–32. doi: 10.1038/s41366-018-0143-9

68. Thompson SV, Hannon BA, An R, Holscher HD. Effects of isolated soluble fiber supplementation on body weight, glycemia, and insulinemia in adults with overweight and obesity: a systematic review and meta-analysis of randomized controlled trials. *Am J Clin Nutr*. (2017) 106:1514–28. doi: 10.3945/ajcn.117.163246

69. Cho SS, Qi L, Fahey GC, Klurfeld DM. Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. *Am J Clin Nutr*. (2013) 98:594–619. doi: 10.3945/ajcn.113.067629

70. Al-Suhaimi EA, Al-Jafary MA. Endocrine roles of vitamin K-dependent-osteocalcin in the relation between bone metabolism and metabolic disorders. *Rev Endocr Metab Disord*. (2020) 21:117–25. doi: 10.1007/s11154-019-09517-9

71. Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, et al. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics*. (2015) 16:493–508. doi: 10.1093/biostatistics/kxu058